A Quantitative Risk Assessment Model for Distribution Cyber Physical System under Cyber Attack

Song Deng, *Member*, *IEEE*, Jiantang Zhang, Di Wu, *Member*, *IEEE*, Yi He, *Member*, *IEEE*, Xiangpeng Xie, *Member*, *IEEE*, Xindong Wu, *Fellow*, *IEEE*

***Abstract*—An accurate and comprehensive risk assessment in a distribution cyber-physical system (DCPS) is essential to ensure its smooth operation, effective control, and exposure of hidden dangers and security implications. Unfortunately, prior risk assessment methods are mostly limited in two aspects. First, the current studies do not respect the complex network topology and intertwined device-wise dependencies, thereby failing to delineate and model the risk propagation mechanism in DCPS accurately. Second, the prior work tends to focus on gauging the risks underlying physical systems independently, overlooking the quantification of risk impact on physical systems incurred by cyber attack. To fill the gap, we propose a unified risk assessment model that gauges the comprehensive security implication of DCPS in a Bayesian regime, in which the three key components, namely, the prior probability, the posterior probability, and the minimum load loss ratio, are calculated via the cumulative densities, the epidemic model, and the optimal load shedding, respectively. We benchmark our proposed model on a public testbed, IEEE 39-bus system, with extensive experiments. The results substantiate the viability and effectiveness of our model and suggest that the vulnerability of DCPS is positively correlated with the three components in our Bayesian modeling. We hope this finding can shed some lights on future research by providing a plausible apparatus for measuring the security implications of physical systems in DCPS under cyber attacks.**

***Index Terms*—distribution cyber physical system, cyber attack, infectious disease model, load loss, risk assessment**

NOMENCLATURE

CDF	Cumulative Distribution Function.
DCPS	Distribution Cyber-Physical System.
DTU	Data Transfer Unit.
ONU	Optical Network Unit.
PP-SIR	Posterior Probability model based on SIR.
RTU	Remote Terminal Unit.
s.t.	Subject to.
SIR	Susceptible-Infected-Recovered.
SVF-EnTop	Solve Vulnerability Factor based on Entropy method and TOPSIS.
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution.
TTU	Transformer Terminal Unit.

This work was supported by the National Natural Science Foundation of China (No.51977113,62120106008,62176070,62022044) . (Corresponding author: Di Wu).

Song Deng is with Institute of Advanced Technology, Nanjing University of Posts & Telecommunications, Nanjing 210003, China (e-mail: dengsong@njupt.edu.cn)

Jiantang Zhang is with College of Automation, Nanjing University of Posts & Telecommunications, Nanjing 210003, China (e-mail: zjt17798569547@163.com)

D. Wu is with the Institute of Artificial Intelligence and Blockchain, Guangzhou University, Guangzhou 510006, Guangdong, China, and also with the Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, China. (e-mail: wudi.cigit@gmail.com)

Yi He is with Old Dominion University, Norfolk, Virginia 23462, USA (e-mail: yihe@cs.odu.edu)

Xiangpeng Xie, is with Institute of Advanced Technology, Nanjing University of Posts & Telecommunications, Nanjing 210003, China (e-mail: xiexiangpeng@njupt.edu.cn)

Xindong Wu is with the Key Laboratory of Knowledge Engineering with Big Data (the Ministry of Education of China), Hefei University of Technology, Hefei 230009 , China (e-mail: xwu@hfut.edu.cn)

I. INTRODUCTION

MODERN distribution networks such as smart grid have emerged to deliver the power (i.e., electricity) from energy resources to end users in an ad-hoc hence highly-efficient manner. To counter against the modeling complexities incurred by the nested network topology and heterogeneous end-devices with multiple voltage levels and diverse security implications, the *control* of a distribution network deeply relies on a cyber system that leverages recent advances in communication and computation technologies including 5G, edge computing, artificial intelligence, to name a few, leading to a distribution cyber-physical system (DCPS).

Despite its promising efficacy and performance, the *robustness* of a DCPS usually remains unclear. The reason is that the interleaved nature of the cyber system and the physical network in DCPS tends to yield a combinatorial large number of flexible access terminals, inviting attackers to perform malicious activities, such as Trojan intrusions [1], sandworm attacks [2], and adversarial machine learning [3]. Indeed, exposure to such attacks may lead to catastrophic failure of a DCPS, where a small set of compromised devices can propagate the malicious information across the entire DCPS through their network dependencies and communications [4]. As a result, an apparatus to gauge the *security risks* in DCPS in a highly accurate means is necessitated and urged. The main challenges of devising such an apparatus are three-fold, described as follows and shown in the middle panel of Fig.1.

1) Complicated risk propagation mechanism -— uncovering the risk propagation mechanism in DCPS is very difficult due to the intertwined topological structures and business processes, frequent interaction, and the coupled functional components of the cyber and physical systems [5].

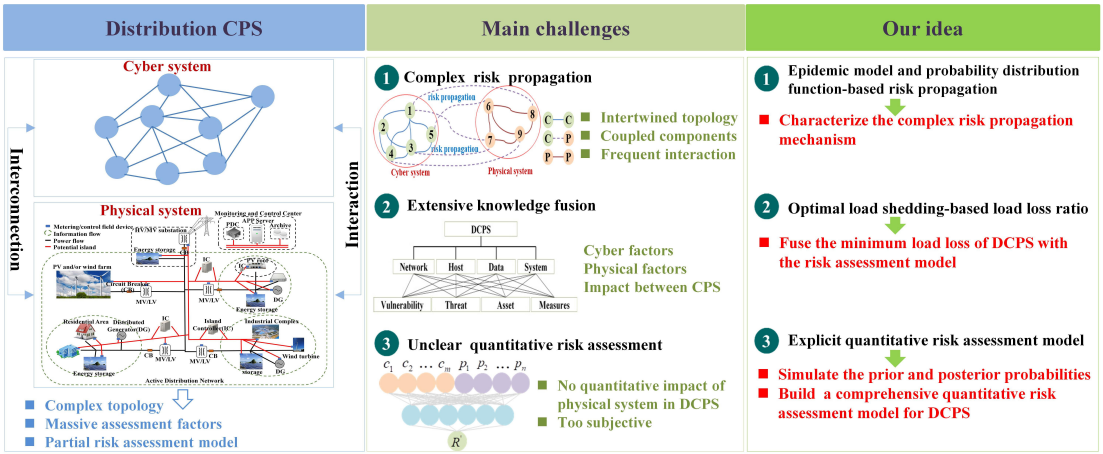


Fig. 1: The three challenges and the corresponding proposed model for quantitative risk assessment of DCPS.

Extensive knowledge fusion -— extensive domain and expert knowledge are required to tailor a risk assessment method for DCPS [6–9], which is laborious, time-costly, and economically inefficient.

2) Unclear quantitative risk assessment — whereas prior efforts have been dedicated to assess the risk impact of physical systems or cyber networks independently [10–13], an assessment of the security implication of the entire DCPS under cyber attacks is left unclear.

In this paper, we deliver a unified quantitative risk evaluation model for DCPS to overcome the three challenges all at once. Its key idea is to frame the quantification of the security risks and network intrusions that irritate the stable operation state of DCPS in a Bayesian regime, which consists of three key components as shown in the right panel of Fig.1. First, an *epidemic model* that simulates the spreading of an infectious disease is employed to characterize the complicated risk propagation mechanism between the complex cyber and physical system in DCPS. Second, the *minimum load loss ratio* of the distribution network after being attacked is quantitatively measured based on optimal load shedding, by which an optimization program constrained by the power flow of the distribution network, voltage, current, and node and branch power balance is solved. Third, the prior probability, posterior probability, and the load loss ratio, in our Bayesian risk assessment model are approximated by three likelihood functions, namely, 1) a DCPS being exposed to cyber attacks, 2) the physical system being attacked conditioned at the succeeded cyber attacks, and 3) the quantitative impact of an attack on a physical system, respectively.

Specific contributions of this paper are as follows:

- A novel quantitative risk assessment model for DCPS is proposed, which does not entail domain and expert knowledge yet is capable of capturing the risk propagation mechanism from the complex network topology and device-wise interactions of DCPS.
- A minimum load loss model is devised to quantitatively gauge the risk impact of the cyber attack on the physical system, which boils down to a constrained optimization program that can be solved via optimal load shedding.
- Extensive experiments are carried out on a benchmark testbed, i.e., the IEEE 39-bus system, and the results substantiate the viability and effectiveness of our proposed risk assessment model.

The remainder of this paper proceeds as follows. Section II reviews the related works. Section III elaborates details of our risk quantitative assessment model of DCPS under cyber attack. Section IV presents experimental results on IEEE 39-Bus and extrapolates findings. We conclude this paper and highlight future directions in Section V.

II. Related Work

In this section, we outline the prior art in the risk assessment models and algorithms in DCPS and discuss the pros and cons of these models and their correlations to our proposal. In a nutshell, the current risk assessment methods can be categorized into the qualitative assessment, quantitative assessment methods, and machine-learning-based assessment methods, with details presented in sequence as follows.

A. Qualitative assessment models

To accurately reflect the security status of the DCPS, researchers have proposed many qualitative assessment models for the security risk calibration. Che et al. proposed a vulnerability assessment model based on Analytic Hierarchy Process (AHP) for urban power grid from four indexes [14], which include power supply, power architecture, operation and transmission channels. Cherdantseva et al. reviewed a variety of cyber security risk assessment methods for SCADA systems, and pointed out the applicable scope, process, and effect of several qualitative assessment methods [15]. Xenias et al. applied the Delphi strategy with domain expert knowledge to the risk assessment of the UK smart grid [16]. *However*, existing qualitative assessment models mainly lack internal links and correlative modeling between different security risk indicators, thereby failing to uncover and delineate the risk propagation mechanism of any DCPS. *In addition*, these methods entail extensive domain and expert knowledge and

use multi-index fusion to assess the security risks of various energy systems, making them laborious and time-costly to be tailored and inaccessible to the users without such knowledge.

B. Quantitative assessment models

Quantitative assessment methods have been proposed for power systems based on mechanism models. Tao et al.[17] proposed a quantitative assessment method based on extinction angle trajectory for commutation failure in power system [18]. Pan et al. proposed a network attack scheme based on the combination of data integrity and availability and formulated a mixed integer linear programming problem by combining vulnerability assessment and attack impact to evaluate the risk of combined attacks on distribution systems [19]. Xiong et al. proposed a dynamic game that evaluates the threat of information network vulnerabilities by considering the allocation of available resources and the cost effectiveness of both sides [20]. Moreover, probabilistic models are also a common practice that fall in this category [21–24]. Although these methods lifted the reliance on domain knowledge by leveraging concrete mathematical modeling, they still fail to reveal the inherent correlation and mutual influence of various security risks in the DCPS.

C. Risk assessment based on machine learning and AI

To better discover and mine the relationship between various security risks in the DCPS, recent advances have been built upon machine learning and artificial intelligence (AI) techiniques. Wu et al. established a risk assessment framework based on improved fuzzy comprehensive assessment method for multi-energy complementary system [25]. Federica et al. proposed a novel online dynamic security assessment model with feature selection [26]. Deb et al. proposed a software-defined network information security risk assessment model based on Pythagorean fuzzy set [27]. Gonzalez et al. proposed a Bayesian attack graph modeling based on which all attack paths were extracted through the attack graph [28]. Chen et al. proposed an elastic statistical risk assessment model based on temporal logistic regression for distribution networks [29]. Wang et al. proposed the dynamic network intrusion detection and prediction model based on fuzzy fractional ordinary differential equation [30]. Zheng et al. proposed a vulnerability assessment method based on deep reinforcement learning in the process of power grid topology optimization under cyber attacks [31]. Wang et al. proposed an improved quantitative evaluation method for attack graphs [32]. Yang et al. proposed a novel risk assessment model based on causality analysis for SCADA system [33]. Unfortunately, these machine learning and AI methods mainly suffer from the lack of explainability, and hence the accountability, in the real-world practices.

D. Remark and Our Thought

The pervasive system vulnerabilities are a main source of power grid security risks, which may exist in any node or branch in a DCPS. By attacking or maliciously manipulating these nodes or branches, a cascading fault can be trigger to propagate erroneous messages throughout the grid, which eventually incurs grid collapse[34]. A highly integrated cyber-physical system in distribution networks further escalates this effect, where security risks can be sourced from the cyber network environment, in addition to the physical grid itself. However, most power-grid security risk assessment models take system vulnerability as the main assessment index, which do not consider cyber security risks. Moreover, the prior security risk assessment models neither accurately reveal the propagation mechanism of security risks between the cyber network and the physical system in the DCPS; Nor do they apply the propagation mechanism and the quantitative loss of the physical system under cyber attack for risk assessment. This paper fills the gap by analyzing the security risks of DCPS through integrating 1) the probability of the cyber system being attacked, 2) the probability of the physical system being attacked conditioned at the succeeded cyber attack, and 3) the minimum load loss ratio of the distribution network after attacks.

III. Quantitative risk assessment for DCPS

In this section, we focus on quantitative risk assessment modeling for DCPS with the prior probability, the posterior probability, and minimum load loss ratio. We note that, although there exists a variety security risk factors from terminal and communication to system in DCPS, an extension to integrate these factors is straightforward. For the sake of self-containedness, a full list of all possible security risks along with the the system architecture are deferred to **S1 in the Supplementary File of this paper**.

A. Quantitative risk assessment model

In contrast to existing risk assessment methods, in this work, we build a quantitative risk assessment model for DCPS from the perspective of prior probability, posterior probability, and minimum load loss ratio. Firstly, a probability model of the cyber system being attacked based on the cumulative distribution function is proposed. Secondly, under the condition that the cyber system is attacked, the probability of the physical system being attacked is calculated based on the infectious disease model. Finally, the minimum load loss ratio of the physical system after being attacked is solved based on the optimal load shedding algorithm. The quantitative risk assessment model for the DCPS is shown in Fig. 2. All risk assessment elements in Fig. 2 come from cyber and physical system in DCPS.

B. Prior probability calculation

The prior probability indicates the probability that an information node tends to be attacked due to the vulnerability of the DCPS. The vulnerability of information nodes not only needs to consider the security protection degree of information nodes, but also the strength of the cyber system being attacked. The higher the vulnerability, the greater the probability of being attacked. Since the cumulative distribution function has the advantage of single parameter and monotonic variation, the cumulative distribution function is chosen in this paper to

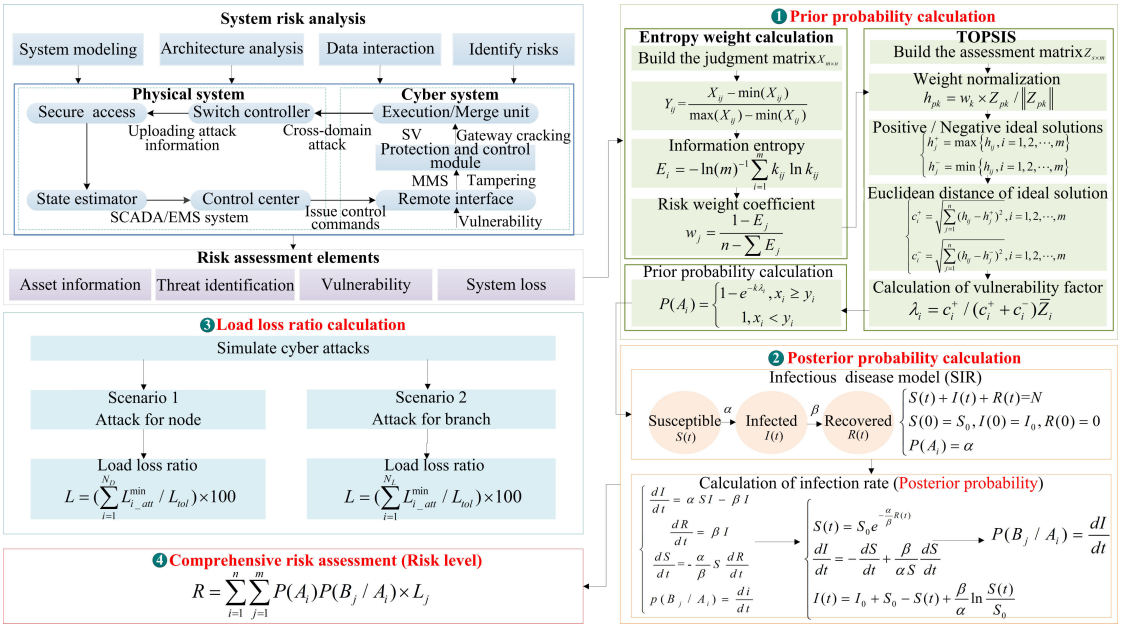


Fig. 2: The quantitative risk assessment model for the DCPS.

represent the effect of the system vulnerability coefficient k on the prior probability in DCPS. Therefore, this paper proposes a prior probability model based on the cumulative distribution function (CDF), presented as follows.

$$P_i = \begin{cases} 1 - e^{-k\lambda_i}, x_i \geq y_i \\ 1, x_i < y_i \end{cases} \tag{1}$$

where λ_i is the vulnerability factor of the *i*-th information node; k is the vulnerability degree coefficient; x_i is the defense strength of the *i*-th information node; y_i is the attack intensity of the *i*-th information node. $P_i = 1$ means that the *i*-th information node cannot resist cyber attacks.

We have the following lemma.

Lemma 1. *$\forall k > 0$, the prior probability of the* i*-th information node in the DCPS satisfies $0 < P_i \leq 1$.*

Proof. For DCPS, the *i*-th information node in the DCPS has security vulnerabilities in the operating system, application software, and communication protocols, so $0 < \lambda_i < 1$. Also, the attack strength y_i and defense strength x_i of the *i*-th information node satisfy $y_i \leq x_i$. Then for $\forall k(k > 0)$, we have $0 < e^{-k\lambda_i} < 1$, and thus $0 < P_i < 1$ according to Eq.(1). Only when $y_i \gg x_i$. Therefore, for $\forall k(k > 0)$, we have $0 < P_i \leq 1$. □

In Eq. (1), λ_i is related to multiple indicators such as network security, system security, and the encryption measures in the DCPS. As a result, the solution of the vulnerability factor belongs to a multi-index decision-making problem. To solve λ_i, we introduce the technique for order preference by similarity to ideal solution (TOPSIS) and information entropy method, and propose an algorithm for solving vulnerability factor based on entropy method and TOPSIS (SVF-EnTop). **This solution is deferred to S2 in the Supplementary File of this paper.**

Definition 1. Let $X = (X_{ij})_{m\times n}, i \in [1, m], j \in [1, n]$, where m denotes the number of security risk indexes in the cyber system, n denotes the number of security risk assessment factors, and X_{ij} denotes the level of the *j*-th risk assessment factor corresponding to the *i*-th security risk index. We call $X_{m\times n}=[X_1, X_2, ..., X_m]^T$ as the judgment matrix of the cyber system.

To calculate the weight $w_i, i \in [1, m]$ corresponding to each security risk index in X, and it is first normalized to obtain $Y_{ij}=\frac{X_{ij}-\min(X_{ij})}{\max(X_{ij})-\min(X_{ij})}$, and then w_i is calculated based on the information entropy.

Definition 2. Let the normalized X be $Y = (Y_{ij})_{m\times n}$, and $k_{ij} = \frac{Y_{ij}}{\sum_{i=1}^{m} Y_{ij}}$, then $E_i = -\ln(m)^{-1} \sum_{i=1}^{m} k_{ij} \ln k_{ij}$ is said to be the information entropy of the *i*-th information node.

According to the entropy weighting method, we can obtain the weight w_i of each security risk index is

$$w_i = \frac{1 - E_i}{m - \sum_{i=1}^{m} E_i}. \tag{2}$$

Definition 3. Let $Z = (Z_{pk})_{s\times m}, p \in [1, s], k \in [1, m]$, where s denotes the number of branches (or nodes) in the DCPS, m denotes the number of security risk indexes in the branches (or nodes), and Z_{pk} denotes the security risk level corresponding to the *k*-th security risk index of the *p*-th branch (or node). Then, $Z_{s\times m} = [Z_1, Z_2, ..., Z_s]^T$ is called the assessment matrix of the DCPS.

Definition 4. Let the assessment matrix of DCPS be $Z = (Z_{pk})_{s\times m}, p \in [1, s], k \in [1, m]$, w_k denote the weight of the *k*-th security risk index, and $h_{pk} = w_k \times \frac{Z_{pk}}{\|Z_{pk}\|}$, then $H = \{h_{pk}\}$ is called the weighted normalization matrix of the DCPS.

Definition 5. Let $H = (h_{pk})_{s\times m}, p \in [1,s], k \in [1,m]$ denote the weighted normalization matrix of the DCPS, $h_k^+ = \max\{h_{pk}\}$ denote the highest level corresponding to the k-th security risk index, and $h_k^- = \min\{h_{pk}\}$ denote the lowest level corresponding to the k-th security risk index. Then H^+, H^- are called the positive ideal solution and negative ideal solution of the DCPS, respectively.

Definition 6. Let c_p^+ denote the Euclidean distance from the p-th information node to the optimal solution in the DCPS, donated as $c_p^+ = \sqrt{\sum_{k=1}^{m}(h_{pk}-h_k^+)^2}$, and c_p^- denote the Euclidean distance from the p-th information node to the worst solution in the DCPS, donated as $c_p^- = \sqrt{\sum_{k=1}^{m}(h_{pk}-h_k^-)^2}$, and $\overline{H}_p$ denote the average of all security risk indexes of the p-th node in the DCPS. Then we call $\lambda_p = \frac{c_p^+}{(c_p^+ + c_p^-)\overline{H}_p}$ the vulnerability factor of information node.

According to Definition 6, we have that $0 \le \lambda_p \le 1$.

After solving the λ_i of the i-th information node in the DCPS, we can get the prior probability P_i that the i-th information node is attacked by bringing λ_i into Eq. (1).

C. Posterior probability calculation

The so-called posterior probability is the probability that other nodes or branches will be attacked under the condition that a node or branch in the DCPS is attacked. Due to the complex topology, diverse communication environments and high integration between cyber and physical systems in DCPS, which in turn makes security risks propagate between cyber and physical systems with the interaction of data flows. This is similar to the propagation mechanism of infectious diseases. To quantitatively analyze the probability of the j-th node or branch being attacked under the condition that the i-th node or branch is attacked in the distribution network, this paper proposes a posterior probability model based on SIR (Susceptible-Infected-Recovered) infectious disease model (PP-SIR).

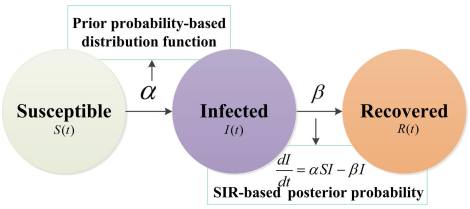


Fig. 3: SIR model.

In Fig.3, α represents the probability of a node or branch being infected, which is related to the vulnerability of DCPS. β represents the probability that the infected node or branch is recovered, and is related to the security tolerance of DCPS. In this paper, we treat the prior probability as α.

Let N be the total number of nodes or branches in the DCPS, and $S(t), I(t), R(t)$ be the number of nodes or branches in the susceptible, infected, and recovered states of the DCPS at time t, respectively. We hold the equation:

$$S(t) + I(t) + R(t) = N. \tag{3}$$

Let $S(t)|_{t=0} = S_0$, $I(t)|_{t=0} = I_0$, $R(t)|_{t=0} = 0$. According to the infectious disease model, the following equation holds.

$$\begin{cases} \frac{dI}{dt} = \alpha SI - \beta I \\ \frac{dS}{dt} = -\frac{\alpha}{\beta} S \frac{dR}{dt} \\ \frac{dR}{dt} = \beta I \end{cases} \tag{4}$$

By solving Eq. (4), we can obtain the following equation

$$\begin{cases} S(t) = S_0 e^{-\frac{\alpha}{\beta}R(t)} \\ \frac{dI}{dt} = -\frac{dS}{dt} + \frac{\beta}{\alpha S}\frac{dS}{dt} \end{cases}. \tag{5}$$

By integrating Eq. (5), the following equation holds

$$I(t) = I_0 + S_0 - S(t) + \frac{\beta}{\alpha} ln \frac{S(t)}{S_0}. \tag{6}$$

Lemma 2. *Let N be the total number of nodes or branches in the DCPS, $S(t)|_{t=0} = S_0 > 0$, $I(t)|_{t=0} = I_0 > 0$, $I_0 + S_0 = N$, $\lim_{t\to+\infty} S(t) > 0$. When $t \to +\infty$, we have $\lim_{t\to+\infty} R(t) = C$, where C is a constant.*

Proof. From Eq. (3), we have $R(t) = N - S(t) - I(t)$, then $\lim_{t\to+\infty} R(t) = \lim_{t\to+\infty}(N - S(t) - I(t))$. Suppose that $\lim_{t\to+\infty} R(t) = \infty$, then according to Eq. (5), we have $\lim_{t\to+\infty} S(t) = \lim_{t\to+\infty} S_0 e^{-\frac{\alpha}{\beta}R(t)} = 0$. However, $ln\frac{S(t)}{S_0}$ in $I(t)$ requires that $\frac{S(t)}{S_0} > 0$. In turn, $\lim_{t\to+\infty} I(t) = \lim_{t\to+\infty}(I_0 + S_0 - S(t) + \frac{\beta}{\alpha} ln\frac{S(t)}{S_0})$ makes no sense. Therefore, it contradicts with the original proposition.

Suppose that $\lim_{t\to+\infty} R(t) = C$, then according to Eq. (5), $\lim_{t\to+\infty} S(t) = \lim_{t\to+\infty} S_0 e^{-\frac{\alpha}{\beta}R(t)} = S_0 e^{-\frac{\alpha}{\beta}C} > 0$, and from Eq. (6), we have $\lim_{t\to+\infty} I(t) = \lim_{t\to+\infty}(I_0 + S_0 - S(t) + \frac{\beta}{\alpha} ln\frac{S(t)}{S_0}) = I_0 + S_0 - S_0 e^{-\frac{\alpha C}{\beta}} - C$. Thus, $\lim_{t\to+\infty} R(t) = \lim_{t\to+\infty}(N - S(t) - I(t)) = N - S_0 - I_0 + C = C$. Therefore, it coincides with the original proposition. □

Remark: From Lemma 2, we know that the nodes or branches infected in the DCPS are repaired with a certain probability as time increases due to various security measures and control policies. Thus, the propagation of the security risk in the nodes or branches of does not cause a complete collapse of the DCPS. From Eq. (4), we know that $\frac{dI}{dt}$ means the change probability of infected nodes or branches under the condition that the susceptible nodes or branches are attacked. Therefore, this paper uses $\frac{dI}{dt}$ as the posterior probability.

D. Minimum load loss ratio calculation

The security threats in DCPS inevitably causes load loss, voltage fluctuations, and power angle instability of the generator set to the physical system of DCPS. Therefore, we integrate the damage degree of the DCPS under cyber attack in the quantitative risk assessment model . In the electric power safety accident regulations, it is clear that the load loss ratio is the standard for dividing the power safety accident level, and the use of load loss to measure the degree of damage to the physical system by network attacks has appeared in a large number of documents. This paper uses the minimum

load loss ratio to quantitatively evaluate the damage of the network attack to the DCPS, and proposes a minimum load loss optimization model based on the optimal load shedding.

For nodes and branches in the DCPS, this paper considers the threat model that covers two attacking scenarios as follows.

Scenario 1. The attacker tampers with the node power through false data injection, causing the node to be overloaded.

Scenario 2. The attacker modifies the circuit breaker opening and closing commands in the branch, causing the attacked branch to be disconnected.

Definition 7. Let N be the number of nodes or branches in the DCPS, $L_{i\_att}^{\min}$ be the minimum reduction of overload load under the condition that *i*-th node or branch is attacked, L_{tol} be the total load of the system in the normal state. Then $L = (\sum_{i=1}^{N} L_{i\_att}^{\min}/L_{tol}) \times 100$ is called the minimum load loss ratio of physical system under the network attack against the node or branch.

The optimal load shedding algorithm is used to solve the minimum load loss of node or branch in the DCPS under cyber attacks. The goal is to solve the minimum load loss ratio that satisfies the power flow, voltage, current, node and branch power balance constraints. This paper takes the calculation of the minimum load loss of a power distribution node under a network attack as an example. The objective function and constraint conditions are described as follows.

$$\sum_{i=1}^{N} L_{i\_att}^{\min} = \min \sum_{i \in N_D} P_i \tag{7}$$

$$s.t. \begin{cases} P_i = U_i^2 g_{ii} + \sum\limits_{j \in \Pi(i)} U_i U_j (g_{ij} \cos\theta_{ij} + b_{ij} \sin\theta_{ij}) \\ Q_i = -b_{ij} U_i^2 - \sum\limits_{j \in \Pi(i)} U_i U_j (b_{ij} \cos\theta_{ij} - g_{ij} \sin\theta_{ij}) \\ U_{i\min} \le U_i \le U_{i\max}, i = 1, 2, \cdots n \\ I_i \le I_{i\max}, i = 1, 2, \cdots l \\ P_{Gi}^{\min} \le P_{Gi} \le P_{Gi}^{\max}, i \in N_G \\ -C_l^{\max} \le C_l \le C_l^{\max}, l \in N_L \\ 0 \le P_j \le P_{dj}, j \in N_D \end{cases} \tag{8}$$

In Eq. (8), P_i and Q_i are the active and reactive power flows of the DCPS, respectively; $\Pi(i)$ is the set of nodes connected to node i; g_{ij} and b_{ij} are the conductance and susceptance of nodes i and j, respectively; U_i and U_j are the voltage amplitudes of nodes i and j, respectively; θ_{ij} is the phase angle difference between nodes i and j; and $U_{i\max}$ are the lower and upper limits of the voltage of node i and n is the number of nodes; $I_{i\max}$ is the upper limit of current of branch i; l is the number of observable branches in the DCPS.

And, N_G, N_D and N_L are the sets of generators, loads, and branches in the DCPS, respectively; P_i represents the balance of input and output power of the generator and load; P_{Gi}, $P_{Gi}^{\max}$ and $P_{Gi}^{\min}$ represent the power of the generator, the upper limit and the lower limit of the power of the generator; P_j and P_{dj} are the actual power and rated power of the node load, respectively.

E. Comprehensive quantitative risk assessment

Based on prior probability, posterior probability and minimum load loss ratio, we build a comprehensive quantitative risk assessment model.

Let $D^{n\times m} \equiv < C^n, P^m >$ denote a DCPS, where $C^n = \{c_1, c_2, ..., c_n\}$ represents the number of information nodes, each c_i is composed of various types of terminals, system software, and application software in the power distribution cyber system, and $P^m = \{p_1, p_2, ..., p_m\}$ represents the number of physical nodes, each p_j is composed of various DTUs, RTUs, TTUs, ONUs, breakers, generators, branches and buses in the power distribution physical system. On this basis, the proposed quantitative risk assessment model of DCPS under the cyber attack takes the form as follows.

$$R := \prod_{i=1}^{n} \prod_{j=1}^{m} P(A_i) P(B_j \mid A_i) \times L_j, \tag{9}$$

where $P(A_i)$ represents the probability that the *i*-th information node in the power distribution is attacked (called the prior probability), $P(B_j \mid A_i)$ represents the probability of the *j*-th physical node being attacked under the condition that the *i*-th information node in the power distribution is attacked (called the posterior probability), and L_j represents the minimum load loss ratio after the *j*-th physical node is destroyed.

IV. Experiments and analysis

In this section, we analyze the risk of DCPS under cyber attack based on the IEEE 39-Bus system to evaluate the feasibility and effectiveness of the proposed model. **The topology of the IEEE 39-Bus system is shown in S3 in the Supplementary File of this paper**. We begin by introducing the experimental settings. Then, we scrutinize the calculation of the prior probability, the posterior probability, and the minimum load loss ratio of the distribution network under cyber attack, respectively. We close this section by presenting and quantifying the risk level of DCPS under cyber attacks based on the experimental results.

A. Experimental settings

All the experiments are implemented based on Matlab 2018b under the Win10 system. The load optimal disaster reduction algorithm in the Matpower 7.0 toolbox is called to solve the optimal load loss. The experimental parameters are shown in Table 1.

TABLE I: Parameters of IEEE 39-Bus

Bus	Parameters				
	The number of generation	The number of node	The number of branch	Total load of generation	Total load of node
39	10	39	46	6297.87MW	6254.23MW

By calculating the prior probability, posterior probability, and the minimum load loss ratio under cyber attack on 39 nodes and 46 branches, the results are brought into Eq. (9) to evaluate the security risk level of each node and branch in the IEEE 39-bus system. However, when some nodes or branches

in the IEEE 39-bus system are attacked, the minimum load loss value is almost negligible. The security risk value of these nodes or branches is almost zero. Therefore, we only select the nodes or branches that pose a greater threat to DCPS security to analyze the security risk level.

B. Prior probability calculation

First, this set of experiments assumes that 1) the vulnerability of DCPS is related to security risk indicators such as the level of defense measures, network environment security, communication protocol security, and system software and hardware vulnerabilities, and 2) the corresponding four evaluation elements are importance, degree of harm, degree of loss, and degree of outbreak. On this basis, we construct the judgment matrix $X = \begin{bmatrix} 4 & 3 & 4 & 4 \\ 4 & 2 & 1 & 1 \\ 4 & 4 & 2 & 5 \\ 3 & 1 & 4 & 4 \end{bmatrix}$, and normalize X. Then, we combine the entropy weight method to obtain the weight coefficient of each security risk indicator. The vulnerability factor λ_i of all nodes and branches is calculated based on TOPSIS. Finally, the prior probability is obtained by substituting λ_i into Eq. (1). Fig.4 shows the vulnerability factor and prior probability of different nodes and branches.

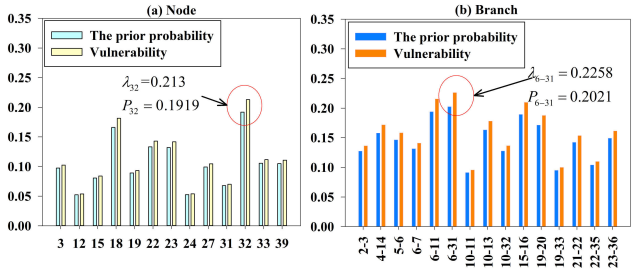


Fig. 4: Vulnerability factor and the prior probability of different node and branch in the IEEE 39-bus system.

Eq. (1) shows that the vulnerability factor is positively correlated with the prior probability. Fig.4 supports that the greater vulnerability corresponds to the greater prior probability. Notably, the vulnerability and prior probability of node 32 and branch 6-31 are the largest, which are 0.1919 and 0.213, 0.2021 and 0.228, respectively. Hence, node 32 and branch 6-31 may have limited security defense measures, uncontrollable network environment, insecure communication protocols, software or hardware vulnerabilities, and other security risks, which makes they be vulnerable. Eventually, their probability of being attacked (i.e., the prior probability) increases.

C. Posterior probability calculation

To calculate the posterior probability, we set $S(0) = 38, I(0) = 1, R(0) = 0$, and $s(0) = 0.97, i(0) = 0.03, r(0) = 1$. The prior probability is the probability that the susceptible state is infected in the infectious disease model. According to the theory of infectious disease model, the probabilities of being infected by different susceptible states (i.e., prior probability α) and being restored to the infected state (i.e., recovery probability β) can affect the proportion of nodes in each state, as shown in Fig.5.

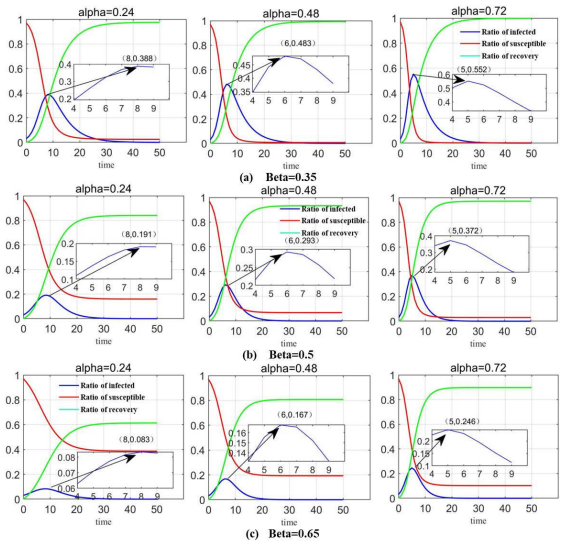


Fig. 5: The proportion of each state node in the IEEE 39-bus system under different α and β.

Fig.5 shows that when the probability of the node being infected is unchanged, our model can maximize the proportion of nodes in an infected state in the shortest time ($t = 5$) with the continuous increase of prior probability. These results verify that a greater vulnerability tends to have a greater prior probability of successful attack, and a greater proportion of nodes in an infected state makes the time of reaching the maximum proportion decrease. Ultimately, the overall risk of DCPS is higher. Therefore, by strengthening DCPS's security protection level and reducing its vulnerability, its probability of being attacked can be significantly reduced.

Besides, Fig.5 also shows that when α does not change, continuous increasing of β can minimize the proportion of nodes in the infected state in the same time, which indicates that a higher β has a stronger recovery ability against cyber attacks. In other words, a smaller prior probability α and a higher recovery probability β are the most beneficial to the security risk defense in DCPS. Therefore, by strengthening DCPS's tolerance and resilience to unknown attacks, its probability of being attacked can be significantly reduced.

Finally, we set α= 0.24, β= 0.65, and substitute them into Eq. (4) and Eq. (5) to calculate the posterior probability of nodes and branches, as shown in Fig.6.

D. Minimum load loss ratio calculation

To simplify the experimental analysis, we does not consider the load loss caused by the different actions of the relay protection device after the DCPS is attacked. In scenario 1, to expand the damage consequences, an attacker will select all nodes to inject false data to overload the system. To ensure the stable operation of DCPS, the system needs to reduce the load, judge the change of the grid load through the power flow calculation, and then calculate the load loss caused by the

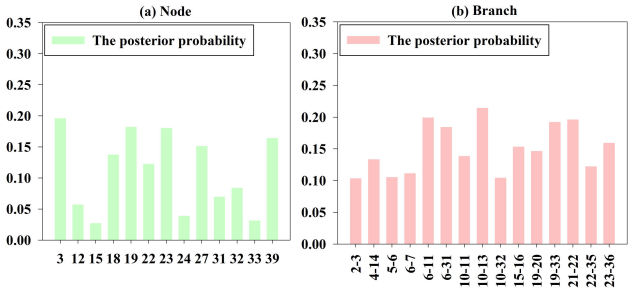


Fig. 6: The posterior probability of different nodes and branches in the IEEE 39-bus system.

attack. In scenario 2, an attacker chooses to attack the branch of DCPS, which causes the load loss of the related nodes. We judge the sum of the load loss of each node after each branch is attacked by enumeration method. Fig.7 shows the load loss of each node and branch under normal and attacked states, respectively.

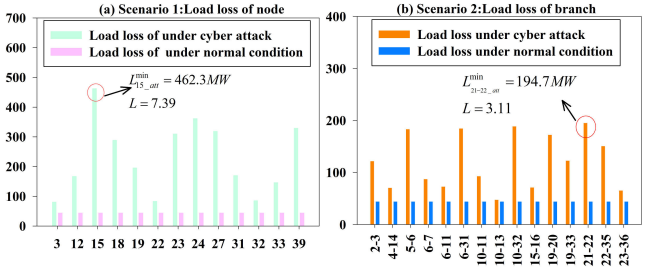


Fig. 7: Load loss of node and branch in the IEEE 39-bus under attack and normal condition.

In scenario 1, Fig.7(a) shows that the load loss of node 15 is the largest. The reason is that node 15 is an important contact node as it is responsible for the power input tasks of generators 33, 34, 35, and 36. Its load loss is 462.3MW, load loss under normal operation is 43.6MW, and minimum load loss ratio is 7.39.

In scenario 2, Fig.7(b) shows that the loss load of branches 21-22 is the largest. The reason is that when node 21 exits operation, the network topology changes and the generator power transmission channel connected to the node 35 is cut off. After branches 21-22 are attacked, the load loss is 194.7MW and the minimum load loss ratio is 3.11.

From scenarios 1 and 2, we find that the data injection attack on each node in scenario 1 makes the system load loss be more serious. The minimum load loss ratio of nodes and branches under cyber attack is shown **in S4 in the Supplementary File of this paper.**

E. Quantitative risk assessment

According to Eq. (9), the security risk value of DCPS is related to the prior probability, the posterior probability, and the minimum load loss ratio under cyber attack. Among them, the posterior probability is related to the prior probability, and the prior probability is related to the multiple of the vulnerability factor (k in Eq. (1)). Hence, we analyze the prior probability under different$k$, as shown in Fig.8. Besides, we analyze the relationship between the minimum load loss ratio and the security risk, as shown in Fig.9.

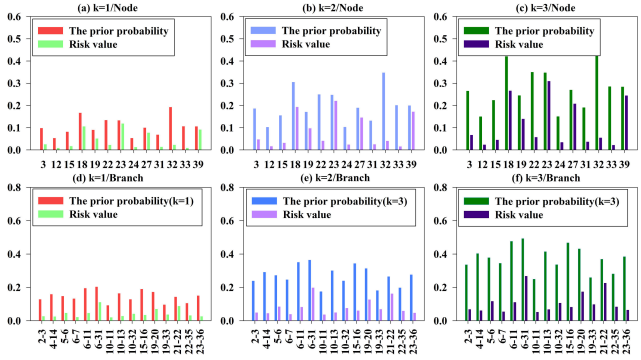


Fig. 8: The security risk value of power distribution under prior probability with different k.

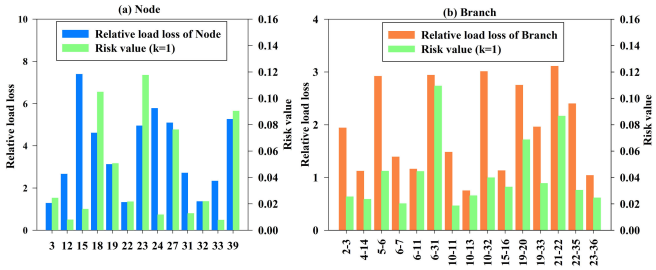


Fig. 9: The relationship between minimum load loss and security risk value of DCPS.

Fig.8 shows that the risk level is closely related to the prior probability. On the same nodes and branches, a larger k increases the prior probability, which ultimately increases the security risk value. Although the minimum load loss ratio of node 15 is the largest (as shown in Fig.9), Fig.8 shows that the risk of node 15 is not the largest because its probability of successful attack is small. Similarly, Fig.9 shows that although the minimum load loss ratio of branch 21-22 is larger than that of branch 6-31, the probability of branch 21-22 of successful attack is less than that of branch 6-31. As a result, the risk value of branch 21-22 is smaller than that of branch 6-31. Therefore, the quantitative risk evaluation is achieved.

V. Conclusions and Future Work

This paper proposes a quantitative risk assessment model for DCPS under cyber attack, where the prior probability, the posterior probability, and the minimum load loss ratio are the three key components. First, the prior probability is calculated based on the cumulative distribution function, entropy method, and TOPSIS. After, the posterior probability is calculated based on an infectious disease model. Finally, the optimal load shedding is employed to calculate the minimum load loss ratio. To evaluate the proposed model, a typical DCPS, i.e., IEEE 39-bus system is employed to conduct the experiments.

The results show that the vulnerability of DCPS is positively correlated with the three key components of the proposed model, which substantiates the viability and effectiveness of our model in quantitatively evaluate the risk of DCPS.

In the future, we plan to take additional evaluation factors, such as the operation mode of the relay protection equipment and the attack of multiple nodes or branches, into account to compile more comprehensive risk assessment studies in the context of robustness analysis of DCPS under cyber attacks.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (No.51977113,62120106008,62176070,62022044).

REFERENCES

[1] C. Liu, P. Cronin, and C. Yang, "Securing cyber-physical systems from hardware trojan collusion," *IEEE Transactions on Emerging Topics in Computing*, vol. 8, no. 3, pp. 655–667, 2017.

[2] A. Farraj, E. Hammad, A. Al Daoud, and D. Kundur, "A game-theoretic analysis of cyber switching attacks and mitigation in smart grid systems," *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 1846–1855, 2015.

[3] J. Wu, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Fault-attention generative probabilistic adversarial autoencoder for machine anomaly detection," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 12, pp. 7479–7488, 2020.

[4] L. Babun, H. Aksu, and A. S. Uluagac, "A system-level behavioral detection framework for compromised cps devices: Smart-grid case," *ACM Transactions on Cyber-Physical Systems*, vol. 4, no. 2, pp. 1–28, 2019.

[5] X. Sun, Y. Liu, and L. Deng, "Reliability assessment of cyber-physical distribution network based on the fault tree," *Renewable Energy*, vol. 155, pp. 1411–1424, 2020.

[6] M. Du and K. Wang, "An sdn-enabled pseudo-honeypot strategy for distributed denial of service attacks in industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 648–657, 2019.

[7] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber–physical system security for the electric power grid," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 210–224, 2011.

[8] D. Choeum and D.-H. Choi, "Vulnerability assessment of conservation voltage reduction to load redistribution attack in unbalanced active distribution networks," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 1, pp. 473–483, 2020.

[9] D. Wu, Y. He, X. Luo, and M. Zhou, "A latent factor analysis-based approach to online sparse streaming feature selection," *IEEE Transactions on Systems, Man, and Cybernetics: Systems(Early Access)*, 2021.

[10] F. H. Jufri, V. Widiputra, and J. Jung, "State-of-the-art review on power grid resilience to extreme weather events: Definitions, frameworks, quantitative assessment methodologies, and enhancement strategies," *Applied energy*, vol. 239, pp. 1049–1065, 2019.

[11] C. Li, Y. Liu *et al.*, "Online dynamic security assessment of wind integrated power system using sdae with svm ensemble boosting learner," *International Journal of Electrical Power & Energy Systems*, vol. 125, p. 106429, 2021.

[12] L. Zhang, W. Bai, H. Xiao, and J. Ren, "Measuring and improving regional energy security: A methodological framework based on both quantitative and qualitative analysis," *Energy*, vol. 227, p. 120534, 2021.

[13] E. Q. Wu, D. Hu, P.-Y. Deng, Z. Tang, Y. Cao, W.-M. Zhang, L.-M. Zhu, and H. Ren, "Nonparametric bayesian prior inducing deep network for automatic detection of cognitive status," *IEEE transactions on cybernetics*, vol. 51, no. 11, pp. 5483–5496, 2020.

[14] Y. Che, J. Jia, Y. Zhao, D. He, and T. Cao, "Vulnerability assessment of urban power grid based on combination evaluation," *Safety science*, vol. 113, pp. 144–153, 2019.

[15] Y. Cherdantseva, P. Burnap, A. Blyth, P. Eden, K. Jones, H. Soulsby, and K. Stoddart, "A review of cyber security risk assessment methods for scada systems," *Computers & security*, vol. 56, pp. 1–27, 2016.

[16] D. Xenias, C. J. Axon, L. Whitmarsh, P. M. Connor, N. Balta-Ozkan, and A. Spence, "Uk smart grid development: An expert assessment of the benefits, pitfalls and functions," *Renewable Energy*, vol. 81, pp. 89–102, 2015.

[17] Q. Tao and Y. Xue, "Quantitative assessment for commutation security based on extinction angle trajectory," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 2, pp. 328–337, 2020.

[18] Z. Wang, G. Chen, L. Liu, and D. J. Hill, "Cascading risk assessment in power-communication interdependent networks," *Physica A: Statistical Mechanics and its Applications*, vol. 540, p. 120496, 2020.

[19] K. Pan, A. Teixeira, M. Cvetkovic, and P. Palensky, "Cyber risk analysis of combined data attacks against power system state estimation," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3044–3056, 2018.

[20] J. Xiong and J. Wu, "Construction of information network vulnerability threat assessment model for cps risk assessment," *Computer communications*, vol. 155, pp. 197–204, 2020.

[21] R. Rocchetta and E. Patelli, "Assessment of power grid vulnerabilities accounting for stochastic loads and model imprecision," *International Journal of Electrical Power & Energy Systems*, vol. 98, pp. 219–232, 2018.

[22] V. Venkataramanan, A. Hahn, and A. Srivastava, "Cp-sam: Cyber-physical security assessment metric for monitoring microgrid resiliency," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1055–1065, 2019.

[23] Y. Xu, M. Korkali, L. Mili, X. Chen, and L. Min, "Risk assessment of rare events in probabilistic power flow via hybrid multi-surrogate method," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1593–1603, 2019.

[24] Y. Zhao, L. Huang, C. Smidts, and Q. Zhu, "Finite-horizon semi-markov game for time-sensitive attack response and probabilistic risk assessment in nuclear power plants," *Reliability Engineering & System Safety*, vol. 201, p. 106878, 2020.

[25] Y. Wu and T. Zhang, "Risk assessment of offshore wave-wind-solar-compressed air energy storage power plant through fuzzy comprehensive evaluation model," *Energy*, vol. 223, p. 120057, 2021.

[26] F. Bellizio, J. L. Cremer, M. Sun, and G. Strbac, "A causality based feature selection approach for data-driven dynamic security assessment," *Electric Power Systems Research*, vol. 201, p. 107537, 2021.

[27] R. Deb and S. Roy, "A software defined network information security risk assessment based on pythagorean fuzzy sets," *Expert Systems with Applications*, p. 115383, 2021.

[28] L. Muñoz-González, D. Sgandurra, M. Barrère, and E. C. Lupu, "Exact inference techniques for the analysis of bayesian attack graphs," *IEEE Transactions on Dependable and Secure Computing*, vol. 16, no. 2, pp. 231–244, 2017.

[29] X. Chen, J. Qiu, L. Reedman, and Z. Y. Dong, "A statistical risk assessment framework for distribution network resilience," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4773–4783, 2019.

[30] Z. Wang, L. Chen, S. Song, P. X. Cong, and Q. Ruan, "Automatic cyber security risk assessment based on fuzzy fractional ordinary differential equations," *Alexandria Engineering Journal*, vol. 59, no. 4, pp. 2725–2731, 2020.

[31] Y. Zheng, Z. Yan, K. Chen, J. Sun, Y. Xu, and Y. Liu, "Vulnerability assessment of deep reinforcement learning models for

power system topology optimization," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3613–623, 2021.

[32] Y. Wang, K. Gao, T. Zhao, and J. Qiu, "Assessing the harmfulness of cascading failures across space in electric cyber-physical system based on improved attack graph," in *Proceedings of the CSEE*, vol. 36, no. 6, 2016, pp. 1490–1499.

[33] L. Yang, X. Cao, and X. Geng, "A novel intelligent assessment method for scada information security risk based on causality analysis," *Cluster Computing*, vol. 22, no. 3, pp. 5491–5503, 2019.

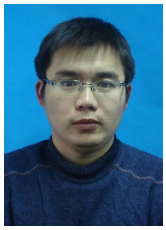
[34] L. Lee and P. Hu, "Vulnerability analysis of cascading dynamics in smart grids under load redistribution attacks," *International Journal of Electrical Power & Energy Systems*, vol. 111, pp. 182–190, 2019.

Song Deng (M'16) received the Ph.D.degree in information network from Nanjing University of Posts and Telecommunication, Nanjing, China, in 2009. From 2009 to 2012, he was a Research Fellow with the State Grid Electric Power Research Institute, Nanjing, China. From 2012 to 2014, he was a Research Fellow with the China Electric Power Research Institute, Beijing, China. He is currently the Associate Professor of Nanjing University of Posts and Telecommunication, Nanjing, China. He was an international visitor with computer science from the University of Louisiana at Lafayette, USA, from 2018 to 2019. His research interests include data security, information security of cyber-physical systems, data mining and knowledge engineering.

Jiantang Zhang received the B.E. degree in electrical engineering and its automation from Sanjiang University, Nanjing, China, in 2019. He is currently pursuing the M.S. degree in computer science at Nanjing University of Posts and Telecommunications, Nanjing, China. His research interests include power grid data security and information security of cyber-physical systems.

Di Wu (M'19) received his Ph.D. degree from the Chongqing Institute of Green and Intelligent Technology (CIGIT), Chinese Academy of Sciences (CAS), China in 2019 and then joined CIGIT, CAS, China where he was an Associate Professor. In 2021, he joined the Institute of Artificial Intelligence and Blockchain, Guangzhou University, Guangzhou 510006, China, as a Associate Professor. He has over 50 publications including journals of IEEE T-NNLS, T-KDE, T-SMC, T-SC, etc., and conferences of ICDM, WWW, IJCAI, ECAI, etc. His research interests include machine learning and data mining. Homepage: https://wuziqiao.github.io/Homepage/

Yi He (S'19-M'21) received his Ph.D. degree in computer science from the University of Louisiana at Lafayette and a B.E. from the Harbin Institute of Technology (China). Dr. Yi He is an assistant professor of Computer Science at ODU. His research focus lies broadly in data mining and machine learning and specifically in online learning, data stream analytics, graph learning, recommender systems, and explainable artificial intelligence. His research outcomes have appeared in premier venues, e.g., AAAI, IJCAI, WWW, ICDM, SDM, T.KDE, T.NNLS, among many other AI, machine learning, and data mining outlets.



Xiangpeng Xie (M'18) received the B.S. and Ph.D. degrees in engineering from Northeastern University, Shenyang, China, in 2004 and 2010, respectively. From 2010 to 2014, he was a Senior Engineer with Metallurgical Corporation of China Ltd. He is currently a Professor with the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, China. His research interests include fuzzy modeling and control synthesis, state estimation, optimization in process industries, and intelligent optimization algorithms. He serves as Associate Editors for the International Journal of Fuzzy Systems, the International Journal of Control, Automation, and Systems.

Xindong Wu (F'11) received the Ph.D.degree in artificial intelligence from The University of Edinburgh, Edinburgh, U.K., in 1993. He is currently a Professor with the Key Laboratory of Knowledge Engineering with Big Data (the Ministry of Education of China), Hefei University of Technology, Hefei 230009 , China. His research interests include data mining and knowledge engineering. Dr. Wu is also a fellow of the Association for American Association for the Advancement of Science (AAAS). He is the editor-in-chief of Knowledge and Information Systems (KAIS).